

March 2022

Design Paper 10

Catalysing responsive and inclusive governance: Using process tracing to build a middle-range theory of governance training interventions

Ryan Sheely, Reva Dhingra and Kamran Hakiman

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

CEDIL design paper

CEDIL design papers present the protocols of studies supported by the CEDIL programme. They describe the objectives, the methodology and the implementation plan of impact evaluations, evidence syntheses, and methodological papers supported by CEDIL.

About this design paper

This design paper was submitted to CEDIL by the “Catalysing Responsive and Inclusive Governance: Developing Theory and Methods for Evaluating Government Training on Citizen Engagement” Project S.163 team.

Suggested citation: Sheely, R., Dhingra, R. & Hakiman, K. (2022) ‘Catalysing responsive and inclusive governance: Using process tracing to build a middle-range theory of governance training interventions’, CEDIL Design Paper 10. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford.

This project was funded by the Centre of Excellence for Development Impact and Learning (CEDIL), supported by UK aid from the UK Government. The views expressed in this research project paper do not necessarily reflect the UK Government's official policies or CEDIL.

All content is the sole responsibility of the authors. CEDIL has not commissioned a professional copy-edit of this paper. Any errors and omissions are the sole responsibility of the authors.



CATALYSING RESPONSIVE AND INCLUSIVE GOVERNANCE: USING PROCESS TRACING TO BUILD A MIDDLE-RANGE THEORY OF GOVERNANCE TRAINING INTERVENTIONS

Design Paper

Center of Excellence for Development Impact and Learning (CEDIL)
Exploratory Study

Ryan Sheely¹, Reva Dhingra², and Kamran Hakiman³

VERSION 2.0, DECEMBER 2020

Abstract

Trainings and capacity strengthening for public officials, citizens, and civil society organizations are some of the most commonly used interventions in Democracy and Governance programming. Despite the prevalence of this type of intervention, **the evidence base of the effectiveness of these interventions is limited**, and **there are no currently existing middle-range theories explaining how and where such interventions do and do not work**. This paper presents the research design for a CEDIL-funded exploratory study that will build a middle-range theory of governance trainings— interventions that involve the use of workshops, courses, or coaching to teach participants about core skills and/or values needed for exercising authority for and on behalf of the public interest. This qualitative study will use **process-tracing case studies** of Mercy Corps' training programs in Myanmar and Jordan to develop a **middle-range theory** that specifies **how**, **where**, and **when** training is effective at changing behaviors and outcomes related to governance. The study will use ongoing, participatory stakeholder analyses with Mercy Corps' program and policy teams in both case study countries and Mercy Corps' London and DC headquarters offices to ensure that the theory and evidence produced by this study are useful and used by practitioners and policymakers.

This design paper provides an overview of the motivation and goal of the study, highlighting the technical design, policy relevance, and methodological innovation of the planned research. It situates the planned study within existing evaluations and theoretical literature, outlines the core methodological approach, and previews the anticipated contributions that the completed study will make to the governance literature, methodology, and policymaking. In particular, this study makes an innovative contribution by building on a set of cutting edge methods for theory-building process tracing to overcome challenges related to developing middle-range theories of change for interventions characterized by extremely weak existing theory and evidence.

¹ Director of Research in Governance and Conflict, Mercy Corps

² PhD Candidate, Department of Government, Harvard University

³ Independent Researcher and Governance Researcher (Consultant), Mercy Corps

1. Background, Motivation, and Goals of Study

1.1 Gaps in Theory and Evidence About Governance Training Interventions

Weak and exclusionary governance is a pressing development challenge, as well as an underlying driver of natural resource degradation, poorly functioning markets, and recurrent cycles of conflict. SDG 16's explicit emphasis on effective, accountable, and inclusive institutions cemented this commitment to treating good governance as a priority development objective. Given this focus by donors, understanding what interventions work to promote good governance is an area for focus for research for academics, donors, and practitioners. A recent [evidence gap map](#) on state-society interactions by 3ie reviews the current base of evidence on governance interventions (Phillips et al 2017). The 3ie review highlights that while the number of impact evaluations on governance interventions is growing, there are still many gaps in coverage in terms of understanding the effectiveness of the full range governance interventions on key outcomes across a variety of contexts.

Three key gaps emerge from existing reviews of evidence on governance (Moehler 2010, Phillips et al 2017, Waddington et al 2019, Justino 2019). First, there are relatively few systematic reviews that aggregate the impacts of specific types of governance interventions. Second, there are relatively few studies that provide evidence about how different types of governance interventions interact with each other. Evidence is particularly weak in terms of understanding how supply-side interventions targeted at formal and informal decision-makers interact with demand-side interventions targeted at communities and/or civil society organizations.⁴ Table 1 provides an overview of the types of supply-side, demand-side, and joint governance interventions identified in existing reviews of evidence.

Finally, the growing body of evaluative research on governance interventions is characterized by a paucity of well-specified middle-range theories that can help to facilitate the transferability of interventions across contexts (Davey et al 2018). In the world of development practice, working theories of governance tend to exist at the level of individual program-level theories of change that provide little opportunity for systematic testing and accumulating findings across contexts (Hakiman and Sheely 2020). Academic research in political science, economics, sociology, and related disciplines helps to define core concepts related to the role of institutions and norms in governance and have advanced a number of overarching theories, but fail to articulate a set of mechanisms and scope conditions about how and under what circumstances development interventions shape the effectiveness, inclusiveness, accountability, and legitimacy of governance processes.

⁴ This distinction between supply-side and demand-side governance interventions follows common usage among governance practitioners, as described in Waddington et al (2019).

Governance Interventions with a Supply-Side Focus	Governance Interventions with a joint Supply-Side/Demand-Side Focus	Governance Interventions with a Demand-Side Focus
Quotas for Increasing Representation of Women and Minorities	Citizen Feedback/Social Accountability Mechanisms	Community Driven Development and Participatory Planning
Training for Politicians and Leaders	Innovations in Democratic Processes and Institutions	Information Dissemination on Service Delivery
Administrative Reform	Electoral monitoring	Civic Education
Performance Incentives	E-Voting	Civil Society Capacity Strengthening
Audits	Decentralization of Decision-making	Negotiation and Alternative Dispute Resolution
	Tax compliance and formalization	
	Land certification	

Table 1. Types of governance interventions included in existing reviews of evidence. Intervention types that predominantly focus on **governance training** highlighted in **bold**. This list of governance intervention types is based on Moehler (2010), Phillips et al (2017), Waddington et al (2019), and Justino (2019). Categorization into supply-side, demand-side, and joint interventions is based on authors' own elaboration.

These three types of gaps in theory and evidence are particularly pronounced for one of the most common types of governance intervention: governance trainings. Broadly speaking, governance training interventions involve the use of workshops, courses, or coaching to teach participants about core skills and/or values needed for exercising authority for and on behalf of the public interest. Governance trainings can be grouped into two broad categories, based on differences in both the target participants and the content of trainings. Supply-side governance trainings focus on politicians, bureaucrats, and informal governance actors such as traditional leaders and armed groups and focus on **the skills and values needed for effective, inclusive, and accountable use of power**. Demand-side governance trainings focus on civic education for community members and/or capacity strengthening for local civil society organizations. These interventions focus on **building the skills and values needed to engage in collective action to influence and monitor decision-makers and directly provide local public goods**. Specific areas of focus for demand-side governance trainings may include voting and civic engagement, advocacy training, and direct participation in citizen feedback and dispute resolution mechanisms.

Both types of governance trainings are often deployed on their own as a standalone program, but are also frequently deployed as part of holistic governance programs alongside other types of intervention described in Table 1. As a result, the amount invested in governance training programs annually is substantial. One study estimates that the annual investment in supply-side trainings alone totals up to US \$15 billion annually (Guy 2016; Denney et al 2017). The scope of demand-side trainings is also massive. USAID-funding alone led to nearly 16 million individuals receiving civic and voter education programs in 2019 (USAID 2019). Likewise, in 2015 the UK allocated £738 million for governance (86% of which flowed through DFID), with a significant portion likely being invested in training (ICAI 2017).

Despite the widespread use of governance training interventions by practitioners and donors, the base of theory and evidence about the effectiveness of this type of intervention is limited. The 3ie evidence gap map identifies that supply side training is one of the least well-studied governance interventions. As of April 2019, only five ongoing/completed impact evaluations assess the effect of such interventions on governance behaviors and outcomes (Phillips et al 2017). The evidence base on demand-side trainings is similarly limited, with findings from a total of three studies indicating that various types of training interventions can shape norms and behavior around conflict management, deference to authority, and political participation (Paluck and Green 2009, Finkel et al 2011, Blattman et al 2014). In addition to this relatively small number of studies, there are also no existing middle-range theories that articulate the key features of both supply- and demand- side governance trainings, the mechanisms by which these trainings shape key behaviors and outcomes, and how training interventions interact with other types of governance interventions and key elements of local contexts. **As noted in the CEDIL inception papers, building better middle-range theories is an essential step towards increasing the use of evidence to inform policy decisions and donor investments (Davey et al 2018). For governance trainings, building a middle-range theory is necessary to accelerate the pipeline of rigorous evaluations and to enhance the transferability and transportability of individual evaluation finding.**

1.2 Goals of the Study and Research Questions

This goal of this study is to develop a **middle-range theory for governance training interventions**. Building on recent literature in philosophy of science, **we define a middle-range theory for a development intervention as an abstract explanation that articulates a set of hypothesized causal pathways connecting an intervention to outcomes and which specifies individual-level and contextual factors that shape how the intervention operates and its expected range of application**. For the sake of clarity, Table 2 (page 5, below) summarizes the definitions of additional key methodological concepts that are used throughout this design paper.

This research project will achieve this core goal by answering the following set of research questions:

1. What are the **causal mechanisms** linking governance trainings to changes in behaviors and outcomes? What is the relative importance of technical knowledge/skill, norms, incentives, and other possible pathways?
2. How are the mechanisms linking governance trainings to changes in behaviors and outcomes shaped by variation in individual and contextual variables (**moderating factors**)?
3. What are the variables that determine the expected range of application (**scope conditions**) on how and where the causal mechanisms underpinning governance training interventions will operate?
4. What types of **evaluation strategies and measurement methods** are needed to test hypotheses following from a middle-range theory about the effects of governance trainings? Which countries and policy domains offer the highest-potential opportunities for rigorously evaluating governance training interventions?

Concept	Working Definition in this Paper	Source(s) of Working Definition	Other Terms Used to Describe the Same Concept
Middle-Range Theory	An abstract explanation that articulates a set of hypothesized causal pathways connecting an intervention to outcomes and which specifies individual and contextual factors that shape how the intervention operates and its expected range of application.	Kaidesoja (2019a), Davey et al (2018)	Middle-Level Theory (Cartwright 2020, Cartwright et al 2020)
Theory-Building Process Tracing	Use of empirical narrative from a case study or set of case studies to articulate a middle-range theory.	Kaidesoja (2019b), Beach and Pedersen (2019)	N/A
Causal-Process-Tracing Theory of Change (pToC)	A visual representation of a Middle-Level Theory of Change for a given type of intervention.	Cartwright (2020), Cartwright et al (2020)	N/A
Causal Mechanisms	A hypothesized causal account of the decisions, behaviors, and interactions of actors that connect a cause to an outcome.	Kaidesoja (2019b), Beach and Pedersen (2019)	Middle-Level Causal Principles (Cartwright 2020, Cartwright et al 2020), Causal Pathways (Sheely and Hakimian 2020)
Scope Conditions	Individual or aggregate-level contextual factors that define limits on where a set of concepts and mechanisms can and can't be transported.	Kaidesoja (2019a)	Expected Range of Application (Cartwright 2020)
Moderating Factors/Variables	A set of individual and/or aggregate-level contextual factors that shape the operation of a given mechanism or mechanisms.	Davey et al (2018)	Moderating Variables (Davey et al 2018), Support Factors (Davey et al 2018, Cartwright 2020), Derailers and Safeguards (Cartwright 2020, Cartwright et al 2020)
Informal Bayesian Reasoning	Using common-language interpretations of the key elements of Bayes's theorem (prior, likelihood ratio, and posterior probability) to structure evaluations of theory and evidence.	Fairfield and Charman (2017), Beach and Pedersen (2019), Zaks (2020)	N/A

Table 2. Working Definitions of Key Research Methods and Approaches

To answer the research questions listed above, we will use **comparative theory-building process-tracing case studies** of programs that Mercy Corps implements in two contexts: Myanmar and Jordan. The two case studies are both typical of contexts where Mercy Corps operates governance trainings— semi-authoritarian and developing states in democratic transition. Myanmar and Jordan differ on important dimensions such that an exploration of the role of moderating factors and scope conditions shaping causal mechanisms will be feasible. In addition, lessons from “shadow cases” of programs implemented by other organizations within these two countries as well as programs implemented in other countries will be applied to extend the generalizability of findings from our primary case studies.

In answering our research questions, we will produce the following:

- 1) **A written account of a middle-range theory** that articulates hypotheses about how, where, and when trainings are effective at changing behavior of both decision-makers and citizens and at improving effectiveness, inclusion, and legitimacy of governance institutions and processes,
- 2) **A causal-process-tracing theory of change** (pToC) that presents a graphical summary of the key elements of the middle-range theory, and
- 3) **A prioritized list of hypotheses and research designs** for rigorous impact evaluations in a set of comparative contexts suggested by the theory.

As we discuss in greater detail in Section 3 below, our set of proposed research methods will help us to answer these research questions and will allow us achieve our goal of building a living middle-range theory that articulates a coherent set of hypothesized mechanisms and moderating factors that can immediately inform program design and which can guide future evaluation and research. In this view, a successful middle-range theory is a starting point for future research and learning that can be updated on an ongoing basis using the results of future studies, rather than a static artifact that will be “finished” upon the completion of our study. As described in Section 3, we will build on cutting edge practice within case study research by developing a set allow researchers, practitioners, and evaluators to identify the relative certainty and uncertainty around each component of our middle-range theory, which will help to transparently ensure that the hypotheses within our middle-range theory can inform future program design, evaluation, and learning. **As we emphasize throughout this paper, the main methodological contribution of this study is building on a set of cutting edge methods for theory-building process tracing to overcome challenges related to developing middle-range theories of change for interventions characterized by extremely weak existing theory and evidence.**

1.3 Overview of Design Paper

Following the guidelines for CEDIL-funded research projects, this paper is designed to serve as the pre-analysis plan for our exploratory study, “Catalysing Responsive and Inclusive Governance,” which is scheduled to run from July 2020 to June 2022.⁵ In particular, the central purpose of this design paper is to highlight the policy relevance, methodological innovation, and technical quality of the proposed research design. Given that the aim of our study is to build a middle-range theory about a type of intervention (governance trainings) for which the existing base of theory and evidence is extremely limited, the logical flow of this paper differs from the structure of sections set out in the design paper guidelines published by CEDIL (CEDIL 2020).⁶

The paper proceeds as follows. In the next section, we situate our planned research within the limited existing empirical literature on governance trainings, as well as several existing bodies of theory that may serve as useful inputs in identifying the mechanisms, moderating factors, and scope conditions that will make up a middle-range theory of governance training interventions. We also briefly discuss the importance of practical approaches such as Thinking and Working Politically (TWP) and Problem Driven Iterative Adaptation (PDIA) to both developing a middle-range theory of governance training interventions and developing a stakeholder engagement plan that ensures that the theory and evidence produced in this study is used and useful. In the subsequent section, we outline the key elements of the technical design of our research study, including defining and elaborating on the proposed methodology of **comparative process-**

⁵ As discussed in the inception report submitted alongside this design paper, these dates are a modification from our original project dates, due to the implementation delays and uncertainty introduced by the COVID-19 pandemic.

⁶ Annex A presents a table indicating how the sections in this paper correspond to the required sections and sub-questions articulated in the CEDIL design paper guidelines.

tracing case studies using **informal Bayesian reasoning**. We next highlight the anticipated contributions of our study in three domains: the substantive literature on governance, methodological innovation, and policy relevance. Finally, we discuss the challenges and opportunities for policy-relevant research generation posed by the COVID-19 pandemic.

2. Relationship to Existing Literature

The body of literature relevant to building a middle-range theory of governance training interventions is substantial, encompassing two main substantive areas: 1) evaluations of governance training interventions and related programming approaches from practitioners and evaluators, 2) academic research on institutions, norms, and democratization in political science, economics, and sociology.

For the purposes of this design paper, we will focus this discussion of the relationship between our study and the existing literature on describing existing evidence on governance trainings. The technical design section of this paper will briefly summarize the key pieces of the methodological literature that have informed our research design and will outline our planned approach for further synthesizing the theoretical and conceptual literature on institutions, norms, and political behavior during the course of our research project. Table 3 outlines the working definitions of key concepts as they are used throughout the course of this design paper. As discussed in sub-section 3.2.2 below, these and other core concepts will continue to be refined throughout the life of the project, and revised definitions will be presented in the final deliverables.

Concept	Working Definition	Source(s) of Working Definition
Governance	The process of exercising political, economic, or social authority.	Risse 2011, Sheely and Hakiman 2020
Good Governance	Exercising political, economic, and/or social authority for and on behalf of the public interest.	Haruna and Kannae 2013, Van Doeveren 2011, Bevir 2012, Sheely and Hakiman 2020
Institutions	Humanly devised constraints that shape human interaction	North 1990, Ostrom 2005, Helmke and Levitsky 2006
Informal Institutions	Rules and procedures created, communicated, and enforced by actors outside of the state or formal organizational structures	Helmke and Levitsky 2006
Governance Norms	A set of shared moral values and principles that collectively define how decision-makers and citizens should behave.	Ostrom 2005, Bicchieri 2017, Sheely and Hakiman 2020

Table 3. Working Definitions of Key Substantive Concepts

2.1 Description of Intervention: Governance Trainings

The concepts of supply-side and demand-side governance are an extension of market theory from economics applied to political systems (Agarwal and Van Wicklin 2012). The categorization delineates between formal and informal supply-side political actors, who are responsible for exercising authority through decision-making, public good provision, or bureaucratic functioning. The corresponding ‘demand-side’ of governance are the constituents, who delegate decision-making and public goods provision to political actors. This terminology is widely used to categorize governance interventions by both academics and practitioners, and is therefore useful for positioning our research for audiences. However, it is worth

noting that a governance actor can fluidly move between ‘demand’ and ‘supply’ side, depending on the circumstances, and simplified accounts of politicians and bureaucrats “supplying” governance and constituents demanding can miss the variety of ways in which political authority is utilized in practice. In practice, there are many types of co-production arrangements that involve direct citizen participation in decision-making, such as community policing or school-based management committees (Moore and Joshi 2004, Levy 2014). These types of arrangements typically involve “demand-side” governance actors playing “supply-side” governance roles.

With respect to governance trainings, the distinction between supply and demand side is used to delineate differences in the intent and design of an intervention by the training provider, not the use of the training materials by the recipient. This is for obvious practical purposes, as information, skills, and beliefs can be deployed in either arena and are generally non-exhaustible. Additionally, the same training program, or even same module, can include components which relate to both types of governance actor (i.e. they can teach a political actor how to hold high-level political actors accountable, while also teaching how to better be accountable to lower-level constituents).

Considering the above discussion, supply-side interventions engage political actors who hold some level of formal or informal authority (politicians, bureaucrats, informal or cultural leaders, and politically engaged civil society organizations) and attempt to impart information, capacity, and beliefs which lead to more effective, inclusive, and accountable uses of power, as they relate to either other political actors or constituents. Conversely, demand-side trainings refer to those interventions which engage political actors and attempt to impart information, capacity, and beliefs which better allow them to successfully request and/or pressure *higher-level* political actors to be more effective, inclusive, and accountable in their use of political power.

2.2 Existing Evidence and Related Literature on Governance Trainings

Direct evidence on the efficacy of governance training interventions is scarce, and the underlying mechanisms and conditions for explaining variation between programs are even less well understood. Below, we consider evidence in three different categories: 1) evidence on training and behavior change among governance actors in development settings, including both supply-side and demand-side governance actors, 2) trainings within multifaceted governance programs, primarily focusing on Community-Driven Development (CDD), and 3) a few promising evidence bases which inform a priori likely mechanisms and moderating factors for how training impacts behavior among governance actors.

2.2.1 Governance Training in Development Settings: Supply-Side and Demand-Side Approaches

Turning first to trainings’ effect on governance behavior, we are only aware of one completed study directly examining the effect of training on supply-side governance actors—broadly understood as political actors who supply public goods. This study, carried out by Baldwin et al (2017), found that providing traditional village leaders with training alone had no detectable effect on normative governance outcomes (such as consultative and inclusive governance). However, when village leaders were trained alongside community leaders, community leaders seemed to act as a check on the power of the chief, making them more consultative in their decision-making (Baldwin et al 2017). This contingent effect demonstrates the importance of an institutionalized accountability structure to ensure that supply-side actors are held accountable to the normative values they are trained in, and that this check is held by a demand-side group.

This study also points out that many of the norms included in the training were not novel to the trainees, and moreover, some attendees complained that the values being advocated were at odds with “local values”, which may also be the case within our planned research contexts in Myanmar and Jordan.

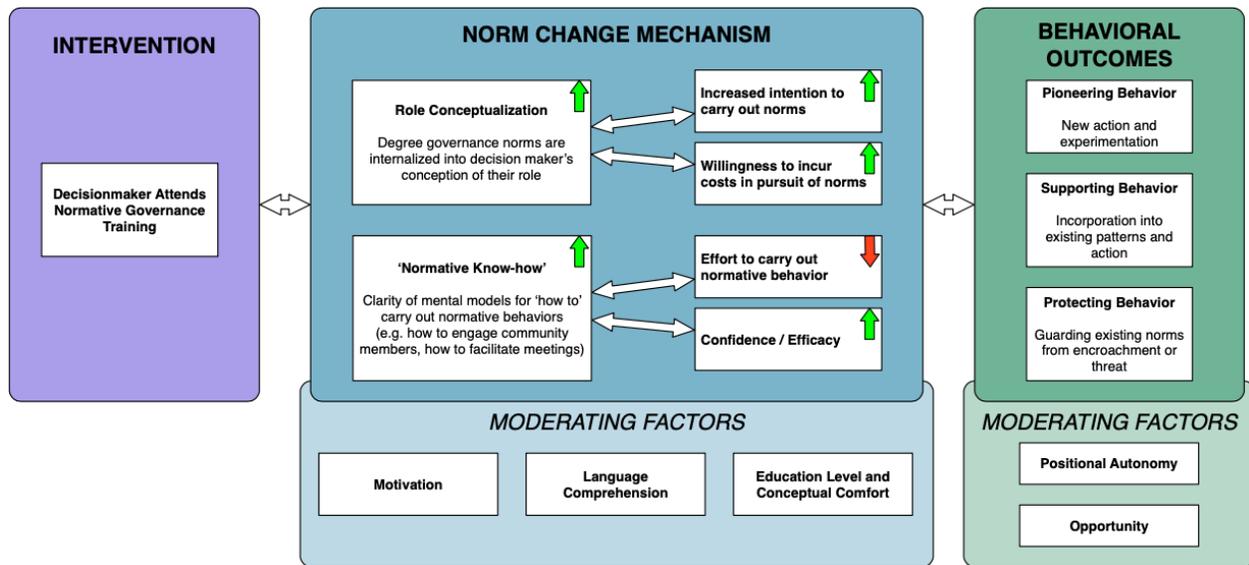


Figure 1. Illustration of Working Theory of How Supply-Side Governance Interventions Shape Norms and Behaviors. Source: Hakiman and Sheely (2020)

In a qualitative study of a Mercy Corps supply-side governance program in Myanmar, Hakiman and Sheely (2020) use a process-tracing case study to articulate two mechanisms through which normative governance training for supply side actors: “*role conceptualization (internalization of norms in relation to their decision-making role) and increased normative know-how (technical knowledge used to better implement a set of values)*. These intermediate mechanisms support the emergence of a wide variety of new behaviors related to good governance, which we group into three broad categories: *supporting (adaptation or improvement of existing activities), protecting (collective action to support rule of law and prevent democratic backsliding), and pioneering (experimentation with new practices that realize good governance principles)*.” This study represents a first step towards building a theory of supply-side governance training interventions by articulating several hypothesized mechanisms and moderating factors (Figure 1). However, this study stops short of articulating a full middle-range theory in several ways. The identified mechanisms only connect the intervention to behaviors by individual decision-makers, without examining how those behaviors shape aggregate patterns of behavior by the broader landscape of supply and demand-side governance. Similarly, this study does not identify mechanisms connecting behaviors and interactions to local level political and service delivery outcomes. In addition, because the study focuses on the analysis of a single program in a single context, it is not able to articulate scope conditions that specify where the hypothesized intervention and causal mechanisms are likely to operate as they did in this case.

The evidence base is wider when considering the behavioral effects of demand-side governance trainings. Like Baldwin et al (2017), these studies demonstrate mixed results, illustrating the need to better understand the formation and effect of norms related to governance. The first, conducted in Liberia by Blattman, Hartman & Blair (2014), found that a training campaign on alternate dispute resolution (ADR) lowered

unresolved conflicts by 29% over slightly more than a year time horizon. Moreover, they find suggestive evidence of behavioral overflow in communities. Blattman, Hartman & Blair (2014) also point out the risk of unintended consequences in governance trainings aimed at shifting behavior. In their experiment, treated villages were more likely to turn to extra-judicial and traditional methods for solving conflict, which they attribute to the training encouraging dispute resolution methods outside the formal state.

Two studies carried out by Steven Finkel and co-authors examine the effect of training at two points in Kenya's governance evolution, and both point to the efficacy of normative democratic training given to citizens. The papers, both focused on Kenya, also examine such trainings in instances of relative democratic ascendancy and democratic backsliding. The first examines governance training during Kenya's democratic transition in 2002, in which Finkel and Smith (2011) find that national training programs were successful in spreading democratic knowledge and values through the National Civic Education Program (NCEP). They find that participants become opinion leaders among their own social groups, further spreading democratic norms and values within society. They also find that participants with less education were more affected by the programs, implying a "compensation effect" (Finkel & Smith 2011). In a subsequent study by Finkel, Horowitz, and Rojo-Mendoza (2012), find promising evidence that normative civic education during democratic backsliding in Kenya's 2007 election, did have a long-term effect on civic competence, though an inconsistent effect on democratic values. It also found that participants later affected by violence were more positive about Kenya's political system and rejected political and ethnic violence—implying that the democratic norms and attitudes imparted were resilient against political shocks, which might lead to democratic backsliding.

2.2.2 Interactions between Training and Other Governance Interventions

Community Driven Development (CDD) programs are perhaps the best documented governance intervention within development. CDD also illustrates that governance trainings elements are often embedded in complex projects, which include tangible outcomes besides the improvement of decision-making in the abstract. This combination is intuitive because tangible programmatic elements provide a substantive focus for practical implementation of the governance skills and norms within the actual community, and sometimes force instances of deliberate contestation, which ideally leads to long-term institutional and normative change (Woolcock et al 2011). Besides providing concrete and context-relevant implementation opportunity, the combination of governance trainings within larger, complex programming also provides a direct incentive (for CDD, this is infrastructure) for a larger range of actors to be interested and engaged in the trainings, extending its potential sphere of influence beyond those intrinsically interested in governance topics.

CDD programs aim to engender good governance through radical decentralization, in which a local institution, usually a Village Development Council (VDC), is created and tasked to implement a local infrastructure project. VDC members are selected through broadly inclusive and democratic means, drawing from the community at large, and carefully designed to include women and minority voices. Relevant for our purpose, CDD programs also include a training component for VDC members.

CDD has been thoroughly evaluated using quantitative and qualitative methods, and the evidence is decidedly mixed. In all cases, however, the role of trainings within CDD is poorly understood, especially the underlying mechanisms for how trainings change the behavior of governance actors. Recent systematic reviews of randomized controlled trials find that CDD interventions have led to marginally better local development outcomes, but find little evidence that these programs lead to long-term improvement in institutions (Casey 2018, White et al 2018). This review considers CDD projects in Indonesian, Afghanistan,

Sudan, Democratic Republic of Congo, Liberia, and the Philippines. The review notes that the ability of CDD to empower minorities and to improve inclusion and participation (especially for women, with the exception of Afghanistan) seems to be either weak or non-durable. Yet, the review is not able to disentangle what elements of CDD make it effective (or not), and what role the training of VDCs had on the ultimate success of the program. This is despite the wide range ratio of funds provided for the facilitation of the training, and the diversity of geographic and social circumstances within which CDD operated.

In relation to CDD, our current study most closely resembles the mixed-method study undertaken by Woolcock et al (2011), in which they outline the complex causal pathways by which CDD facilitates the challenging of elite preferences by the community, and the gradual shifting of governance norms. Though the study deftly addresses certain moderating factors and mechanisms, it does not address the role of training as a separate component.

2.2.3 Power, Political Settlements, and Doing Development Differently: Barriers to Reform and Implications for Practice

An additional relevant body of theory and practical writing in international development highlights how the configurations of power and coalitions among actors underpin institutions and constrain the available space for institutional change (Acemoglu and Robinson 2008, North et al 2009, Khan 2010, Levy 2014, Pritchett et al 2017). This body of theory focuses on how the nature and organization of power between actors and coalitions shapes the operation of institutions and the implementation of policies.

In recent years, this body of theorizing on political settlements has surpassed earlier research on institutions in terms of influence on development practice and policymaking, grouped broadly under the heading of Doing Development Differently (DDD) (Bain et al 2016, Booth et al 2016, Fritz 2017, Wild et al 2017, Honig and Gulrajani 2018). This theory-informed shift in practice can be grouped into two related approaches: 1) Problem-Driven Iterative Adaptation (PDIA) and related approaches toward adaptive and emergent management and capacity building (Andrews et al 2017) and 2) Thinking and Working Politically (TWP) and related approaches to integrating political economy analysis into program design and evaluation (Booth 2015, Kelsall 2016).

This body of theory and practice has important implications for building a practically-oriented middle-range theory of governance training interventions. The types of elite interests and actions highlighted in the political settlements literature suggest a potential set of causal mechanisms and moderating factors that explain how and under what circumstances elite actors may block or co-opt governance training programs and when they may champion and support them. Recent research has indicated that despite the rising popularity of DDD approaches, donors and practitioners have faced substantial challenges with respect to systematically integrating political analysis into adaptive approaches to programming and integrating iterative problem-solving into democracy support and governance as a technical area (Kashwan et al 2018, Menocal and Domingo 2018). Second, understanding the interests and power undergirding policymaking is also emerging as an important factor explaining success and failure in evidence use in international development (Pritchett 2002). Given these two ways that the political settlements and DDD literatures are relevant to our planned studies, these approaches will inform both our theory-building case studies as well as our use of stakeholder analysis and adaptive management in the evidence use plan required by CEDIL.

3. Technical Research Design

There are numerous quantitative and qualitative research designs and methods that can be used to develop policy-relevant middle range theories. The selection of methods to use when developing middle-range theories depends in large part on the existing body of theory and evidence relevant to a given type of intervention. As described in the review of the literature above, governance training interventions are characterized by a **lack of middle-range theories** and a **weak base of evidence from rigorous impact evaluations**. Our proposed research design will attempt to respond to these dual gaps by using a novel combination of qualitative methods to develop an initial middle-range theory that is sufficiently detailed to inform future impact evaluations, facilitate cumulative aggregation of findings, and support the transportability and transferability of evaluation findings to other contexts.

In particular, our research design makes a novel methodological contribution in its use of **theory-building process tracing** with **informal Bayesian reasoning** in **nested qualitative case studies** of a set of **similar programmatic interventions** that are being implemented in **multiple contexts**. While each of these methodological approaches are not individually sufficient for building a middle-range theory, in combination they are able to jointly provide leverage that can help with theory building for interventions where the existing base of theory and evidence is weak. This section presents a walkthrough of the core methods and steps in data collection that will be used in our proposed technical approach, with attention to the standard of rigor and limitations of the proposed qualitative methods. Table 2 in the introduction presents the working definitions of several of the key methods and approaches that will be used in the study.

3.1 Overarching Goal: Building a Middle-Range Theory

The objective of this study is to develop a middle-range theory with the following components (building on definitions of middle-range theories in Davey et al 2018 and Kaidesoja 2019a):

- Conceptualization and definition of a set of **causes** and **outcomes**,
- A set of **hypothesized mechanisms** that link the causes and outcomes,
- A set of individual or aggregate-level **moderating factors** that are hypothesized to shape the operation of a given mechanism or mechanisms,
- A set of individual or aggregate-level **scope conditions** that define limits on where a set of concepts and mechanisms can and can't be transported.

As described in Section 1.2, the middle-range theory produced in this study will be presented in a written narrative that clearly and succinctly describes each of the above components. We will also present the middle-range theory as a causal-process tracing Theory of Change (pToC), drawing on the recent CEDIL methods brief and working paper by Nancy Cartwright and colleagues (Cartwright 2020, Cartwright et al 2020). We will use our stakeholder engagement plan to further identify the most effective ways to present the middle-range theories to different audiences for this research.

The following discussion of methods focuses on explaining why the proposed selection of methods is well-tailored to the purpose of building a Middle-Range Theory for interventions with **limited existing theory and evidence**. Our selection of methods and their application to the country cases (Myanmar and Jordan) are driven by 1) selecting the methods best suited to answering our primary research questions in a theory-building context with limited existing evidence; 2) combining widely used existing qualitative methods and an innovative, quantitatively informed Bayesian approach to produce a

methodological exemplar for other MRT building projects; and 3) producing a final MRT that can be useful for and intelligible by **academic audiences, donors, and implementing organizations**.

In the following sections, we explain and justify how theory-building process tracing and informal Bayesian reasoning can be used to develop empirically-grounded hypotheses about the mechanisms connecting interventions to outcomes and the interactions between these mechanisms and aspects of the local context. Building out a Bayesian approach to developing a MRT and pToC is also meant to bridge the divide between theory development and theory updating, and allow implementers in diverse contexts, including contexts beyond Myanmar and Jordan, to continue developing and refining the core theory of change, alternatives pathways, moderators, and scope conditions. Finally, in making decisions about methods, we are guided by the need for our methods to be replicable in practice by other researchers and across the different contexts in which Mercy Corps and other implementers provide governance trainings. A more detailed description of qualitative data collection and analysis methods and their standards of rigor is available in Annex B.

3.2 Building a Middle-Range Theory: Methods and Rationale

3.2.1 Theory-Building Process Tracing

Process tracing is a key methodological tool for case-based qualitative research used in both theory generation and theory-testing. Theory-building process tracing requires 1) intensive engagement with existing literature to identify possible mechanisms; 2) defining the universe of cases to be explained by a theory; 3) selecting suitable cases for theory generation; and 4) compiling data—interviews, documents, and secondary sources—to identify diagnostic evidence in favor or against theorized mechanisms or raising alternate potential mechanisms.

Following Beach and Pedersen (2019), we first use reviews of existing programmatic, theoretical, and empirical literature to define key concepts and identify causal mechanisms identified in other bodies of theory that could be relevant in building a middle-range theory of governance training interventions. The purpose of this method is to “theorize causes and outcomes in ways that are compatible with mechanistic explanations” and to “look at existing theorization for a source of inspiration for processes for which to search” (Beach and Pedersen 2019, 271). In theory-building process tracing, this set of methods is used to help guide research by suggesting priority respondents and interview questions but will stop short of formulating testable propositions and systematically evaluating them using confirming and disconfirming evidence (Beach and Pedersen 2019).

When undertaking theory-building process tracing of a development program, defining causes entails defining the intervention or interventions that are being studied, as well as disaggregating the working parts of those interventions. Before beginning fieldwork, we will use existing academic and programming literature to build on the working definitions of supply-side and demand-side governance trainings developed above and to further specify the core components of each kind of intervention. We will also use program documents and interviews with Mercy Corps program staff and partners in Myanmar and Jordan to identify which governance training components are present in the respective programs. At this time we will also identify the other types of interventions deployed by Mercy Corps alongside governance training within the same program. Within this, we will work to articulate the causal mechanisms which training specifically is meant to effect and how this relates to the other components of the overall program. For clarity, we will state

reasonable counterfactuals for how the program would be different if the training was not implemented, and how this would affect other mechanisms and outcomes. During the fieldwork and analysis, we will also gather observations on how the intervention components were actually implemented in practice to develop as accurate of a picture of possible of what the intervention in each case actually was.

Defining outcomes and hypothesized mechanisms will follow a similar process. Before beginning fieldwork, we will build on the literature review in this design paper to identify and define the measured and hypothesized outcomes of governance training interventions. We will also review the proposals and monitoring and evaluation plans for each of the focus programs to identify the outcomes and theories of change/logic models that each program specified. This review will be used to generate a list of hypothesized outcomes and mechanisms before the start of interviews with program participants in each of the two research sites. While this list of hypothesized outcomes and mechanisms will be used to help identify specific field sites, interviewees and to draft interview guides, we also expect that it is very likely that fieldwork will allow us to identify both outcomes and mechanisms that were not previously specified in existing literature or program documentation.

This set of steps of using desk research and interviews with program teams to define and theorize interventions, outcomes, and candidate mechanisms will be undertaken jointly by the three US-based researchers and the field research fellows based in Jordan and Myanmar. Specific literature review, document review, and background interview tasks will be divided among the researchers. These desk research tasks will also include background research necessary to complete the more detailed stakeholder analysis that will inform the program's thinking and working politically approach and theory/evidence use strategy. Each researcher will bring draft definitions and candidate mechanisms identified during their review to the full research team, which will then arrive at a shared set of working definitions and candidate mechanisms through a process of collective deliberation and consensus.⁷

3.2.2 Bayesian Informal Reasoning in Process Tracing

Bayesian reasoning is the intuitive idea underlying most research: namely, that new empirical evidence should marginally update our beliefs in a logically consistent manner. Zaks (2020) succinctly states that a Bayesian analytical framework allows researchers to meaningfully answer the question, “*What is the probability that our main hypothesis (H_M) is correct, given that we searched for and found evidence E_i ?*” Despite its recent popularity, the application of Bayesian reasoning within case study research is still relatively new. Zaks goes on to point out unresolved questions regarding Bayesian process tracing's capacity to truly provide an iterative, temporally independent approach, which ‘fully engages rival explanations’ and counteracts the biases of confirmation bias and ad hoc hypothesizing – especially in a manner which is efficient enough to account of its considerable increase in effort and therefore opportunity cost (Zaks 2020).

Our research plan aims to grapple with these questions through building an MRT related to governance trainings in a specific context, not of pure research, but in understanding how a Bayesian approach can be operationalized to build a MRT in a manner which can then be usefully updated by implementing organizations (most notably Mercy Corps but also other peer organizations) as implementing organizations. This mandate of **usefulness for an implementing organization** is closely related to the goal of research but distinct. The evaluation units at Mercy Corps and other implementing organizations are tasked with providing evidence about the effectiveness of an intervention, but currently there is no system by which to accumulate evidence related to implemented ToCs, even when the implemented programs rely on similar program structures, content, and assumed mechanisms. Our research contribution is based on the idea that,

⁷ As needed/possible, Mercy Corps program and policy team members in Myanmar and Jordan may also be included in the desk review and collective process of setting priors around the definition of core concepts and identification of causal mechanisms.

in the case when an organization’s prime mandate is to implement programs based on the best available evidence, there is great value for providing a light-weight and intuitive structure into which the organization can place its generated evidence, by which it can gradually update its ToC in a logically consistent manner. We argue that Bayesian process tracing is uniquely well suited for this task.

In exploring how to operationalize a Bayesian approach, the research team grappled with the limitations inherent in complex implementing organizations, and a prime limitation is variation of technical ability between countries and contexts. Thus, we have decided to pursue an **informal approach to Bayesian thinking**, which, in essence, is a heuristic starting point for identifying priors and likelihood ratios derived from beliefs which can be stated in common language (Table 4). This common-language approach will allow multiple actors to engage in the identification of priors, while an overly formal approach for deriving these priors would exclude stakeholders from meaningfully engaging in the construction and potential contestation of these assumptions, which is one of the purported strengths of Bayesian process tracing.

Linguistic form	Numerical equivalent	Prior/posterior confidence	Likelihood ratio
Certainly, no question about	1.0 (100%)	Higher levels, more confident that mechanism present based on existing knowledge.	Level of certainty (higher levels, more certain to find)
Almost certainly, beyond reasonable doubt	0.9 (90%)		
Very probably	0.8 (80%)		
Probably	0.7 (70%)		
On balance, somewhat more likely than not	0.6 (60%)		
Like as not, even money	0.5 (50%)	Lower levels, less confidence that mechanism present based on existing knowledge.	Level of uniqueness (lower levels, more unique if found)
Somewhat less than even chance	0.4 (40%)		
Probably not	0.3 (30%)		
Very probably not	0.2 (20%)		
Almost certainly not	0.1 (10%)		
Certainly not, impossible	0.0 (0%)		

Table 4. Numerical Probabilities Expressed in Linguistic Form, recreated from Beach and Pedersen, 2019

Our research will begin with an empirics-first approach, as described by Beach and Pedersen (2019), in which we will observe governance training being implemented in Myanmar and Jordan. Using this empirical information, alongside an intensive desk review of related empirical and theoretical work, we will articulate a

relatively naïve MRT, which may deviate substantially from the stated ToC for each Mercy Corps program. Each researcher (three US-based and two research fellows will independently create these a visualized MRT, alongside). We will use these naïve MRTs as a starting point for identifying multiple possible causal pathways, and the research team will decide in practice – and depending on how unique the different MRTs are – how to balance the need for consolidation and simplicity with the value of allowing alternate hypothesized mechanisms to be considered independently and continue to be updated alongside ‘most likely’ pathways.

After articulating this list of definitions and pre-existing mechanisms, the research team will collectively utilize Bayesian informal reasoning in each researcher will independently articulate their prior confidence ($p(h)$) in three aspects of the list: 1) the likelihood of observing the intervention components as described in the empirical record of each case, 2) the likelihood of observing the hypothesized outcomes in the empirical record of each case, 3) the likely operation of each hypothesized mechanism in each case. For any hypothesized mechanisms identified in the literature, researchers will also identify “observable implications” of that mechanism, and for each observable implication will articulate natural language formulations of the certainty ($p(e|h)$)—the likelihood of observing the predicted evidence if the hypothesized mechanism is operating) and uniqueness ($p(e|\sim h)$)— the likelihood that the predicted evidence would be generated if the hypothesized mechanism is not operating in the case) (Beach and Pedersen 2019). This process of articulation of priors, certainty, and uniqueness will use natural language to describe levels of confidence, following standards used in other approaches to using informal Bayesian reasoning in qualitative research (Zlotnick 1972, Beach and Pedersen 2019, Chapter 5).

Below we provide a highly stylized “zoomed in” exemplar of a representation of components of a MRT as a pToC, based on our previous research in Myanmar, with placeholder certainty values in common language.⁸ This graphic only represents one step in a potential MRT. Refer to Figure 1 above for the expanded pToC representing the broader set of hypotheses produced in our previous study (Sheely and Hakiman 2020).

⁸ Probabilities for each mechanism and moderating factor are indicative. This graphic only represents one set of links in the MRT/pToC; downstream connections to outcomes and alternate pathways not pictured.

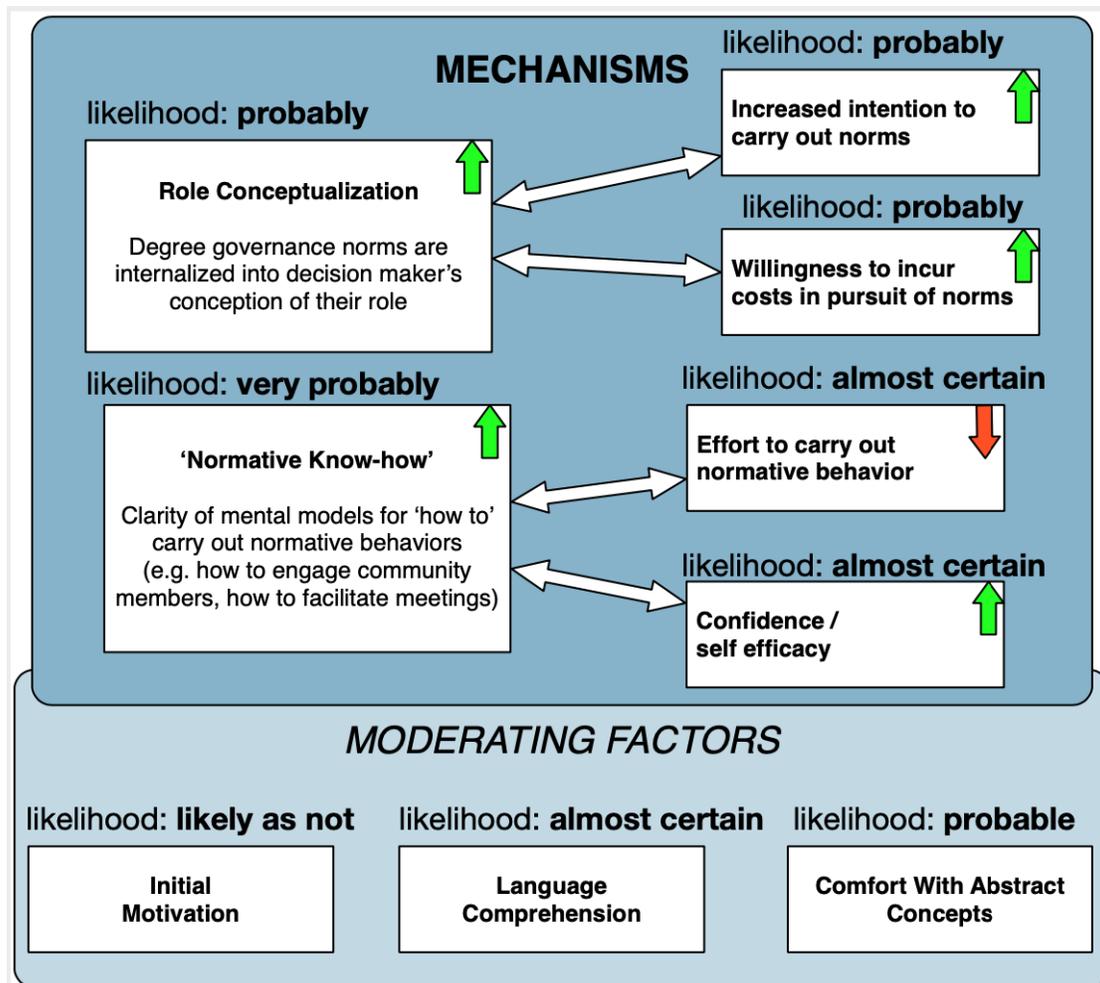


Figure 2. Exemplar of a segment of pToC summarizing a middle-range theory of governance training interventions and natural language representations of probabilities. Source: Hakimian and Sheely (2020)

Though we state probabilities using common language equivalents, the actual application of this language *will* be formalized using the underlying mathematical translation. A key methodological design challenge is developing an approach that can potentially encompass multiple causal mechanisms while meaningfully parsing how these mechanisms relate to different expected outcomes (or empirical fingerprints, as described in Beach and Pedersen). Through testing and iterating during this research, we aim to develop practically oriented methodological tools for how to clearly specify outcome measures, and then relate those to one-or-more of the mechanisms posited in the working MRT.

Note that we do not pre-commit to a specific approach to this. We will begin with the approach outlined in Beach and Pedersen, but adapt that approach iteratively, as the projects confront the complexity of empirical evidence on the ground and can more accurately discern the technical capabilities of implementing teams, which will determine what 'practical' outputs look like.

Though our project is theory **building**, we expect certain methodological outputs to be oriented towards theory **updating**, for use by program and evaluation teams. Our project does not assume that policy-makers or practitioners are interested in theory development for its own sake. With regards to our own theory generation, we **do not** expect programming or policy audiences to merely continue the work of this research (as this is both outside their mandates and dedicated resources). Instead, our tools will anticipate the updating of program theories of change **between** different programs and the iterative implementation of

programs in the same and different contexts. Particularly, we are interested in providing more than a simple narrative MRT in relation to programming and policy. Bayesian reasoning is particularly suited for encompassing multiple, alternative pathways and prioritizing which pathways seem *more* or *less likely* based on information. These means that de-prioritized pathways may become relatively *more likely* based on the updating of the prioritized pathways in our MRT, and ultimately shift the programming and advocacy positions of Mercy Corps and other implementing and funding organizations.

There are several possible limitations to using this type of group-based application of informal Bayesian reasoning to define the core interventions and outcomes and identify candidate mechanisms. First, this approach is time-intensive relative to standard approaches to desk review and concept formation, which typically involve less deliberation and documentation of priors. Second, power dynamics within the research team could lead to convergence of definitions and priors around specific researchers, rather than allowing the diversity of the research team to be harnessed as a resource for improving the creativity and context-sensitivity of the middle-range theory. Third, because this method includes a focus on identifying preliminary hypotheses about possible mechanisms, there is a risk of becoming overly fixated on testing those hypotheses, rather than using them as an input into an essentially abductive process of theory-building (Beach and Pedersen 2019). All three of these limitations will be addressed through careful management and facilitation of this process and through the development of facilitation guides and tools that the research team will be able to use in this process of identifying definitions and candidate mechanisms. Moreover, in this case a desk review would not be suitable given the sparse empirical research on governance trainings, especially supply-side trainings. As such, a Bayesian approach—which demands iterative data-input and explicitly provides space for testing competing/alternative hypotheses—would be impossible through secondary research.

Finally, for this analytical application to be compelling to a wider audience within Mercy Corps' program and policy team members (both in-country and HQ-based), the research team will also need to articulate and engage with the interests and power of the broader national and international stakeholders connected to the project. Most importantly, the research team will need to be able to defend the prior probabilities assigned to different stakeholders, to ensure that the resulting analysis is compelling, and also deal with potential disagreements. This process will also help identify additional interviewees outside of the list of direct program participants, and will also help to develop a politically smart plan for preliminary stakeholder engagement necessary for successful implementation of the research and eventual uptake of the evidence and middle-range theory produced in this study (Levy and Walton 2013, Booth 2013, Kelsall 2016, Andrews et al 2017).

3.2.3 Rationale for Selection of Case Study Contexts and Programs

Appropriate case selection is a core element of case study research (Gerring 2004, Gerring 2007, Seawright and Gerring 2008). For case study research on a development program, case selection involves selecting both a country context and a specific program.⁹ Following guidance for case selection in theory-building process tracing (Beach and Pedersen 2019), we focus on selecting cases that are typical of the types of contexts where governance training programs are implemented and programs that are representative of the type of intervention that is the focus of the theory building (Seawright and Gerring 2008). At the same time, we use the principles of diverse case selection to ensure some cross-case variation in contextual and program-design factors, as this variation will help to identify the types of moderating factors and scope conditions that are a core component of middle-range theories (Seawright and Gerring 2008, Beach and Pedersen 2019). Finally, given that this study is being led by the research team at Mercy Corps, case selection was driven by the organization's country presence and programming portfolio.

⁹ As we discuss in more detail below, we also use a nested case study design, in which mini-case narratives of specific individuals, departments, or localities will be used alongside the overall narrative of each program.

Jordan and Myanmar were selected as the contexts for our theory-building case studies, as both are typical cases of how and where Mercy Corps has implemented governance training programs that also have key political and institutional differences that shape Mercy Corps' programs. Jordan is a monarchy that has attempted democratization reforms over the past decades with mixed success. The recent influx of refugees has meant that two in ten residents are either refugees or migrant workers, placing enormous strain on the country's resources. Myanmar has been grappling with a series of primarily ethnic conflicts since independence in 1948. Despite the progress made in the transition toward a more democratic form of governance, challenges remain around local governance processes, and communities in both countries still have relatively few opportunities to voice their priorities for development. These institutional and political features are compounded by crucial economic differences. While Jordan's GDP per capita nearly four times that of Myanmar, Jordan also receives over 15 times as much net ODA per capita as Myanmar, which is consequential for programs that aim to boost state capacity

Within these two contexts, the theory-building case studies will focus on two programs: the CSSF-funded Leadership and Community Development program in Jordan and the USAID-funded Advancing Community Empowerment program in Myanmar (Table 5). As demonstrated in the table, the respective training components which form the basis of our case studies in each country are fairly similar in form. In both cases, trainings are being provided to local-level officials (including both formal governance and civil society officials), meant to increase their ability and motivation to carry out inclusive governance processes. In the case of Jordan, this training is related to strengthening citizen voice and upward advocacy by local CSOs and local governance actors, including community and youth leaders. In contrast, Myanmar's focus is on the planning, budgeting, and provision of public services through inclusive community engagement. In both cases, the CATALYSE approach (a community engagement curriculum developed by Mercy Corps and implemented internationally) has been a key framework undergirding the design of these training components, though the underlying toolkit and field materials have been significantly adapted to meet the needs of the program and context (Gurung et al 2017). This process of adaptation and associated rationale for the adaptation process in each context will be an additional point of inquiry during qualitative data collection for each study.

Leadership and Community Development Program		Advancing Community Empowerment (ACE)
Country/Region	Jordan 18 Communities All Over the Country	Myanmar 7 States and Regions
Donor	UK-CSSF	USAID
Partners	Community Based Organizations (CBOs) in Each Community	Pact, Mercy Corps, Community Partners International (CPI), and Save the Children International (SCI)
Program Dates	2016- 2021	2017- 2022
Objectives	Building capacity of local government representatives, civil society representatives and community leaders to identify and respond to sources of tension between Jordanian Communities and Syrian Refugees.	Improving the ability of local governance actors to inclusively plan for and implement the provision of public services.
Governance Training Components	Community Engagement for Governments Curriculum; Interest-Based Negotiation Training	Inclusive community engagement for elected leaders and government administrators on public service provision; Interest-Based Negotiation.
Other Program Components	Community-Driven Development Projects, Supporting National Network of Community Leaders (registered as a new CSO), Community Social Events	Community-based dispute resolution training and advocacy training provided to CSOs and community actors.

Table 5. Summary of Case Study Programs

While this approach to case selection is largely in line with methodological best practices, a major limitation is that the selection of typical but diverse cases was constrained by access to Mercy Corps programs, rather than a systematic mapping of all contexts with governance training interventions, and all implementers working within those contexts (Lieberman 2005). As a result, while our research is well positioned to identify the core components of a middle range theory of governance training interventions, these two programs and contexts may not encapsulate the full range of variation needed to identify all possible moderating factors and scope conditions. We will address these limitations in two ways.

First, where possible, we will develop brief “shadow cases” of other programs and contexts that help to develop and refine hypotheses about mechanisms, moderating factors, and scope conditions. These shadow cases may include programs implemented by other organizations in our case study contexts (which could be built on existing qualitative data collection plans) or programs in other country contexts (which would draw on desk research and remote interviewing). The review of shadow cases will enable us to better achieve the goal of developing a middle-range theory of the causal mechanisms undergirding governance trainings by examining the conditions under which the mechanisms identified apply to other contexts (Cartwright et. al 2020). Second, as described in more detail below, when articulating the middle-range theory and writing up final deliverables, we will return to questions of case selection, highlighting what aspects of the theory we were and were not able to specify using the case studies of programs in Jordan and Myanmar. In doing this, our middle range theory and related write-ups will help to suggest case selection criteria for further theory testing and theory building.

4. Anticipated Contributions of Study

As described above, the primary aim of “Catalysing Responsive and Inclusive Governance” is to start to fill the substantial gaps in theory and evidence about governance training interventions by using process tracing case studies of Mercy Corps programs in Jordan and Myanmar to build a middle-range theory. Successful completion of this goal throughout the life of this exploratory study will allow us to make three broad kinds of contribution: 1) to existing literature on governance interventions, 2) to methodological innovation in the evaluation and research communities, and 3) to the policy and practical aims of CEDL, FCDO, and the broader field of international development practice. We briefly address each of these anticipated contributions in turn.

4.1 Contributions to Existing Literature

Based on the review of existing literature above, we anticipate that this study will make contributions to four main bodies of literature.

First, this study will contribute to the nascent evaluation literature on governance training interventions and on governance interventions more broadly (Phillips et al 2017, Justino 2019). In particular, the middle-range theory that we will develop will help to systematize and interpret the small number of existing evaluations and to create shared knowledge about the most pressing theory and evidence gaps. This will help to create coherence in a nascent research field and will assist with knowledge accumulation and collective learning as the number of rigorous evaluations across contexts begins to grow.

Second, this study will contribute to the broader academic literature focused on how institutions and norms change by helping to link the mechanisms underpinning the effectiveness of discrete development interventions to the types of social, political, and organizational process that lead to broader institutional change (North 1990, Ostrom 2005, Boesen 2007, Mahoney and Thelen 2009, Bicchieri 2016).

Third, this study will contribute to the practically oriented literatures on Doing Development Differently, Problem-Driven Iterative Adaptation, and Thinking and Working Politically (Stein and Valters 2012, Booth 2015, Kelsall 2016, Andrews et al 2017, Honig and Gulrajani 2018) by incorporating a focus on configurations of stakeholder interest, power, and coalitions into both the middle-range theory that we are building and in our plan for ensuring stakeholder analysis and evidence use. We expect that using a political economy lens in both the theory-building and stakeholder engagement activities will reveal ways in which similar underlying power dynamics and coalitions shape both the implementation of governance training

programs and broader evidence-use and uptake. Combining these two modes of analysis will also suggest how governance trainings (and evidence about these types of interventions) can be designed and implemented in politically-smart ways that increase stakeholder buy-in and support (Kelsall 2016). In this way, our study will help to overcome the paradoxical challenges that democracy and governance interventions have faced with respect to integrating Thinking and Working Politically and political economy analysis into program design and learning (Kashwan et al 2018, Menocal and Domingo 2018).

In addition, our proposed methodology will make a contribution to the practice of PDIA (and related approaches), by allowing teams to explicitly apply informal Bayesian reasoning to articulate hypotheses underpinning a set of program design choices and transparently update their assessments and pivot in a theoretically grounded, transparent manner (Woolcock 2013, Pritchett et al 2013, Andrews et al 2017). PDIA relies on an implicitly Bayesian framework, in which iterative cycles of action and reflection are meant to lead an implementing group towards a better solution for a problem by updating the solution based on experience. However, PDIA does not provide clear guidance on how to update an intervention once implementation experience has been gained. Besides relying on general consultation and adaptation, a structured, if non-formalized (i.e. non mathematical) approach to updating priors in relation to potential *competing* priors may provide useful guidance and structure to problem-focused teams working in governments across the globes (Andrews et al 2017).

Finally, as we describe in more detail below, this study will contribute to the methodological literature on process tracing in political science, evaluation, and related disciplines (George and Bennett 2005, Woolcock 2013, Befani and Stedman-Bryce 2017, Fairfield and Charman 2017, Fairfield and Charman 2019, Beach and Pedersen 2019, Zaks 2020). While there is a rich literature on the use of process tracing in the social sciences, including in Beach and Pedersen (2019)'s volume on the subject, the debate on the use of informal Bayesian reasoning in process tracing approaches remains fruitful ground for innovation. By both articulating a rigorous research design for the use of our methods and providing an example of its effective application in the context of MRT building, our study is a crucial addition to the methodological literature.

4.2 Methodological Innovation

4.2.1 Description of Methodological Innovation

The methodological innovation of our study is its use of **theory-building process tracing with informal Bayesian reasoning** in **nested qualitative case studies** of a set of **similar programmatic interventions** to develop a **middle-range theory**. Our main contribution is taking a set of existing qualitative data collection and analysis methods and combining them in a new way that is calibrated to the specific methodological challenge we are facing: building a policy-relevant middle-range theory for a type of intervention that is characterized by severe gaps in both theory and evidence.

While the innovation and rigor of our proposed combination of methods are discussed in extensive detail in Section 3, it is worth highlighting several elements of our anticipated methodological contribution in plain language. First, our study makes a contribution by clearly defining and listing the elements that should be specified in a middle-range theory. While the existing literature is successful in distinguishing middle-range theories from other types of theories and articulating underlying epistemological assumptions, there are relatively few clear guides that researchers and practitioners can use when building middle-range theories and using them to inform programming and policy (Davey et al 2018, Kaidesoja 2019b).

Second, our study is innovative in specifying a set of practices and tools that teams of researchers, evaluators, implementers, and/or advocates can use when applying informal Bayesian reasoning to theory-building process tracing (Beach and Pedersen 2019, Kaidesoja 2019b). While the literature on Bayesian process tracing has expanded recently in recent years, it is predominantly focused theory-testing or in advocating such quick iterations of theory-testing and theory-building that the distinction disappears (Befani and Stedman-Bryce 2017, Beach and Pedersen 2019, Zaks 2020). Indeed, the Beach and Pedersen volume, which is one of the most extensive methodological guides for process tracing only devotes 15 of nearly 300 pages to theory-building process tracing and refer to theory-building as an “iterative and creative process” at multiple points (Beach and Pedersen 2019). Our study will build on this approach to explain in more detail how their approach to Bayesian reasoning in process-tracing case studies can be deployed in the context of building a middle-range theory focused on a development intervention with a weak pre-existing base of theory and evidence. In addition, our approach is innovative in that it will produce a set of facilitation guides and tools that large, diverse, cross-cultural teams can use to harness the logic of Bayesian reasoning when developing and updating empirically-grounded middle-range theories.

Finally, our study is innovative in providing an approach for using comparative analysis of multiple process-tracing case studies to develop two key elements of middle-range theories: hypotheses about moderating factors and scope conditions. Most existing treatments of process-tracing focus primarily on within-case analysis in a single case (Gerring 2004, George and Bennett 2005, Checkel 2006, Waldner 2012). Newer studies that describe the comparative use of case studies do not explicitly focus on building middle-range theories or the use of such methods in the context of learning about the effectiveness of development programs (Guala 2010, Koss 2015, Bengtsson and Ruonavaara 2017). Our study builds on these existing approaches to comparative process tracing by articulating a set of methods and tools for extending Bayesian reasoning to making cross-case inferences about the operation of mechanisms, moderating factors, and scope conditions.

4.2.2 Contribution of This Study’s Methodological Innovation to CEDIL’s Research Agenda/Programmes of Work

The methodological innovation in this study will contribute to CEDIL’s research agenda by helping to advance the “enhancing evidence transferability” programme of work. The aim and design of this study is directly inspired by questions and debates raised in CEDIL inception paper on the role of Middle-Range Theories enabling evidence of the effectiveness of interventions to be transferred to other contexts (Davey et al 2018). In particular, it is our expectation that the methodological approaches and tools that we will develop throughout the course of this study will be particularly helpful in jump-starting middle-range theory development, rigorous evaluation, and evidence transfer for interventions and sectors that are characterised by substantial gaps in both theory and evidence. As described in section 1.1 above, governance training interventions (and the broader field of governance practice) are characterized by exactly this kind of dual theory and evidence gaps. This middle-range theory that we will develop in this study will help to promote the generation and transferability of evidence about governance interventions by specifying a set of hypothesized mechanisms that explain how such interventions work, as well as hypotheses about the contextual factors that moderate the operation of those mechanisms and set bounds on where the mechanism can be expected to operate.

4.2.3 Methodological Exemplar

The methodological innovation described in this study will be tested through the process of implementing the proposed process-tracing case studies of governance programs in Myanmar and Jordan. We will use

tools and techniques for ongoing monitoring, reflection, and adaptation from PDIA (and other approaches to adaptive learning/management) to critically assess the performance of the tools and make any needed course corrections throughout the life of the project. At the conclusion of the project, we will compile and edit a consolidated set of lessons learned from this piloting process, along with methodological templates and user's guides for using deliberative Bayesian reasoning in theory-building process tracing and will submit these tools with our final set of deliverables to CEDIL. Where helpful and appropriate, we will further disseminate examples of our methodological innovation in action by contributing to newsletters, blogs, and/or lesson learning papers.

4.3 Policy Relevance

As described throughout this design paper, the direct aim of this study is to generate useful and used theory and evidence about governance training interventions. In the short run, **the case studies of governance training interventions in two challenging and important contexts for governance practitioners: Jordan and Myanmar.** The case study evidence and corresponding middle-range theory will inform discussions about the design of training in both contexts, as well as debates about the role of training alongside other policy, programming, and diplomatic efforts in both countries.

In the long-run, the broader contribution of this study to evidence use will be through the use of the **middle-range theory of governance training interventions** to help organize the relatively small amount of existing evaluation evidence in this area and to **fill the substantial evidence gaps in this sector** by helping donors and implementers to prioritize key evaluation questions, methodologies, and contexts by providing a practically usable structure with which to accumulate evidence, which relies on Bayesian logic. This is an oft cited gap in the practitioner space — where evaluations are donor facing, and inconsistently and indirectly inform programming and policy/advocacy work.

At its core, our study aims to provide an alternative for Mercy Corps' internal evaluation, programming, and policy/advocacy teams, in which the evidence required for evaluations—especially of governance programming involving trainings—will be able to be accumulated in a logical and structured format. This will allow program teams to directly update their working hypotheses, and making explicit the process by which prior beliefs in efficacy and generalizability of program models are made explicit and tested, representing a marked shift, both technically and philosophically, structure for how evaluations relate to underlying theories of change within Mercy Corps' work structure.

Furthermore, this approach is meant to encourage shifting the design and usage of evaluation data towards being theoretical relevance, leading to a natural synergy between research and evaluation (which are separate teams and functions within Mercy Corps). For policy and advocacy teams in Mercy Corps and beyond, this process will provide a solid basis for defending policy positions and a logical rationale for why policy stances might change (i.e. as alternate pathways / explanations become more-or-less likely based on Bayesian updating). Most importantly, this process of cumulative learning will allow Mercy Corps and others to support reform champions in governments, civil society, and communities to effectively advocate and build coalitions in support of interventions that are effective at transforming institutions and norms.

As discussed throughout this paper (and our inception report and stakeholder engagement/evidence use plan), this study is well positioned to be useful for policymakers, practitioners, evaluators, and researchers for several reasons. First, our research team is composed of a set of practically-engaged researchers from Jordan, Myanmar, and the United States who collectively have a unique mix of methodological training and

policy-relevant experience. In addition, our research team’s collective networks span a wide array of networks of academics, practitioners, and evaluators across Jordan, Myanmar, the UK, and the US.

In addition, as described throughout this design paper, this study will use ongoing participatory stakeholder analysis to inform both the substance of our middle-range theory and our evidence use plan. By integrating stakeholder analysis into the core of our project and ensuring that these analyses draw on the collective wisdom and insights of the full range of team members at Mercy Corps throughout the lifespan of our project, we will aim to develop theory and evidence that helps to avoid the trap of promising governance interventions that fail due to an inability to actually take local patterns of politics and power seriously (Andrews 2013, Levy 2014, Honig 2018).

Finally, our ability to produce useful and used evidence will be enhanced by our ability to draw on Mercy Corps’ substantial policy and advocacy expertise and reputation in both of the case study countries and with relevant donors. These country-office and HQ policy colleagues will help us to craft and frame our research questions and findings in ways that respond to pressing policy debates and will be able to use their networks and relationships to help deliver our policy-oriented products to the right stakeholders. In order to ensure that the products of this study are useful to Mercy Corps programming and advocacy colleagues, we will check-in quarterly with two existing organisation-wide working groups: the Governance Practitioners Group (which is made up of governance-focused field practitioners from Mercy Corps’ global field team) and the Peace, Conflict, and Governance Influence Matrix team (which is comprised of HQ-based Advocacy and Programming colleagues). These practitioner groups were involved in the the initial proposal for this research study and have expressed interest in a MRT of governance training interventions, for the purposes of adapting program designs to local contexts, identifying and prioritizing learning opportunities, and making more targeted policy recommendations about the types of contexts where governance training programs are most likely to be effective.

4.4 Implications of COVID-19

The COVID-19 pandemic has ravaged communities across borders and has already posed a serious challenge to the progress of this research project. As such, the research team has both capitalized on the original design features of the study and integrated new elements to ensure that research is able to continue in a safe manner, mitigate risk and further disruption, and contribute to policy discussions brought to the fore during the pandemic.

The design of the study includes research conducted by both the research team in the US and in Jordan and Myanmar. Given the restrictions on in-person research posed by Harvard’s IRB and local regulations in Myanmar, the study design has been shifted to be entirely virtual for the foreseeable future. However, the location of the two research fellows in-country and strong communications between these fellows, the Mercy Corps program teams, and the US-based principal investigators will enable research to continue even in the face of further lockdowns that might disrupt Mercy Corps trainings. For example, research fellows will be able to access program documents and conduct interviews via phone or video call in-country without restrictions posed by international calling or time zone coordination facing the US-based research team.

The pandemic has also both undoubtedly influenced the programs under study and underscored the importance of research on governance trainings. The recent parliamentary elections in Jordan, for example, highlight a gulf between citizens and local leaders that has only expanded with the pressures posed by the pandemic on services and employment. As citizens demand more of their elected officials in increasingly

dire situations, understanding how international programming aimed at improving governance actually achieves these goals is more crucial than ever. The study will include such considerations in both the implementation of research and production of the final academic and policy outputs.

Likewise for Myanmar, the COVID-19 pandemic has stressed the ability of local governance actors to retain public trust, where local elected leaders (many of whom will be involved in the program that is the focus of our case study) were widely blamed for mismanaging a means-tested transfer program in response to the pandemic. The government has been fairly successful at keeping infection rates low, however this has come at enormous economic cost, and the next few years will be critical as the country navigates the pandemic-created recession alongside its continued efforts at a democratic transition. As in Jordan, we will include such considerations in both the implementation of research and production of the final academic and policy outputs.

Annex A: Mapping Contents of Design Paper to Sections Outlined in CEDIL’s Guidance Note

CEDIL Guidance Note Section	CEDIL Guidance Note Subsection	Section/Subsection in Design Paper	Page(s) in Design Paper
Policy Relevance	Description of Intervention	1.1 Theory and Evidence Gaps About Governance Training Interventions	Pages 2-4
		2.1 Description of Governance Trainings	Pages 6-7
Policy Relevance	Goals of the Study	1.2 Goals of Study and Research Questions	Page 4
Policy Relevance	Contribution to Existing Literature	2.2 Existing Evidence and Related Literature on Governance Trainings	Pages 7-12
		4.1 Contributions to Existing Literature	Pages 23-24
Policy Relevance	Relevance for CEDIL, FCDO, and Evaluation/Research Community	4.3 Policy Relevance	Page 26
Innovation	Methodological Innovation	4.2 Methodological Innovation	Pages 24-25
Innovation	Contribution to CEDIL’s Research Agenda	4.2 Methodological Innovation	Page 25
Innovation	Methodological Exemplar	4.2 Methodological Innovation	Pages 25-26
Technical Design	Illustration of How Intervention is Supposed to Work	2.1 Description of Governance Trainings	Pages 6-7
Technical Design	Description of Study Design and Analysis Methods	3.2 Using Multi-Case Process Tracing to Build a Middle-Range Theory: Research Design and Methods	Pages 13-23

Annex B: Detailed Description of Qualitative Data Collection and Analysis Methods

B.1 Qualitative Data Collection in Each Case Study Program/Context

In each case study program/context, a variety of **qualitative data collection methods** will be used to collect the empirical material that will be used for analysis and theory-building. The potential qualitative data collection methods will include in-depth semi-structured interviews, focus groups, observation of ongoing program activities, and analysis of additional program documents and training materials (Yin 2015, Chapters 6 and 7). For each case study context, the working list of data collection activities, sampling strategies for target participants, and draft questionnaires will be jointly informed by the desk review focused on concepts and mechanisms, as well as an up-to-date assessment of the types of data collection (in person vs. remote) activities that are feasible given the COVID-19 situation in each country at the outset of the planned data collection period.

The major potential limitation of this set of qualitative data collection methods is inconsistent or ad hoc usage of the methods within or across the two case study contexts. To address this limitation, In parallel, the research team will convene a series of training sessions on the core data collection methods, to ensure consistency across researchers and to allow for opportunities for mutual learning. While the entire research team has extensive experience implementing qualitative data collection and analysis methods, each researcher has distinct methodological training and strengths. As a result, each member of the research team (including the field research fellows) will take turns leading training sessions focused on methodological best practices for each of the qualitative data collection methods that will be used in the study.¹⁰

Once tool design and training is complete, data collection will begin, with the field research fellows taking the lead on conducting in-person and remote data collection. Each field research fellow will do a weekly reflection and debrief call with at least one of the US-based researchers. When remote interviews are used as a COVID-19 adaptation, the US-based researchers may also join or lead select interviews. The first round of qualitative data collection will be focused on **individual-level narrative and experiences** of governance training program participants and non-participants before, during, and after the program. This round of data collection will also focus on interviews with a broader set of **stakeholders related to the implementation of governance training programs** in each study context, which will be used to further deepen the initial stakeholder analysis that informed our evidence use plan.

The second round of qualitative data collection will focus on **identifying and investigating potential mini-cases** of particular individuals, organizations, localities, or events that illustrate important instances of the presence or absence of hypothesized mechanisms connecting the program to the outcomes of interest (Sheely and Hakiman 2020). In this round, qualitative data collection will be focused on focused **constructing and triangulating the empirical narrative** within each mini-case,

¹⁰ We will use a variety of sources as key reference texts for developing our data collection guides and trainings. For semi-structured interviews, see Diefenbach (2009), Newcomer et al (2015), Gentles (2015), and Kallio et al (2016). For focus group discussions, see Longhurst (2003), Ivanoff and Hultberg (2006), Parker and Tritter (2006), Liamputtong (2011), and Nyumba et al (2018). On participant observation and “rapid ethnography”, see Yanow (2003), Schatz (2013), Gillespie and Michelson (2011), Wessells et al (2012), Jamshed (2014), and Shah (2017).

which will involve repeated in-depth interviews with initial interviewees, as well as interviews or focus groups with additional individuals connected to a given mini-case.

An additional limitation of this approach to qualitative research methods in the context of this study is potential bias that can be introduced due to the fact Mercy Corps is both the lead research agency for this study and the implementing organization for the programs that will be the focus of our case studies. This can lead to multiple possible sources of bias. First, Mercy Corps team members (from both field offices and headquarters) may be inclined to present overly favorable interpretations of program implementation and effectiveness, which could lead to suggestions of respondents or documents that present the programs in the best possible light. This limitation will be mitigated in part by the independent status of Mercy Corps's research team relative to program teams at headquarters and in country offices. In addition, this study's explicit, deliberative use of Bayesian reasoning creates an opportunity for Mercy Corps' team members to clearly state their prior beliefs about a program's underlying mechanisms and outcomes and to bring those priors into dialogue with the evidence.

Finally, while the independent members of our research team are engaged as pro-bono consultants with Mercy Corps for the life of this research project, their outside perspective will help to encourage the Mercy Corps research, program, and policy teams to question unspoken assumptions and challenge biases and blind spots about our own programs. Similarly, the field research fellows will be hired as consultants and managed most closely on a day-to-day basis by the US-based research rather than Mercy Corps programming or policy staff. We will use this management structure to help ensure that the qualitative data that the data collected by the field research fellows represent the full range of perspectives and experiences with the target programs, rather than just observations that confirm Mercy Corps' team members prior beliefs.

Second, Mercy Corps' joint role in implementing this research and the programs that are the focus of the research may lead respondents to provide answers that they think Mercy Corps wants to hear. The independence of the research team vis-a-vis the program teams will also help to mitigate this limitation. In all data collection exercises, the field research fellow (or US-based researcher) will introduce themselves as an independent researcher affiliated with Mercy Corps, and will emphasize their independence from the formal program team during the informed consent statement. Where possible, the field research fellows will use repeated engagement with interviewees to build the trust and rapport necessary to encourage answers that reflect the respondent's lived experiences and perceptions.

B.2 Within-case Qualitative Analysis

The full set of data collected within each case study context will be analyzed using a set of **qualitative data analysis methods**: 1) coding, 2) identifying hypothesized mechanisms and moderating factors, and 3) stakeholder analysis (Yin 2015, Chapters 8 and 9, George and Bennett 2005, Beach and Pedersen 2019, Bryson 2004). The first qualitative analysis method is **coding** of interview and focus group notes/transcripts, program documents, and narrative write-ups of mini-cases identified during the data collection process (Richards and Morse, 2012, Blair 2015, Saldana 2015, Belotto 2018, Maher et al 2018) . We will utilize an iterative, multi-stage coding process using NVivo, in which we will begin with low-level descriptive thematic codes, which we will then group into larger conceptual buckets as larger patterns and connections emerge (Burnard 1991, Hsieh and Shannon 2005, Elo and Kyngas 2008, Baralt 2012). One of the main validity risks with coding qualitative data is inter-coder reliability (Belotto 2018). We will address this challenge by holding a second set of training sessions with the entire research team, focused on establishing a shared set of approaches for coding and qualitative

analysis. In addition, qualitative coding for each case will be led by two members of the research team, with a third team member helping to spot check and cross-validate coding.

Once the qualitative data has been coded, the research team working on each case study will use those codes to **critically examine the core concepts** at use in the study and to identify **hypothesized mechanisms** that may be operating in the case (George and Bennett 2005, Beach and Pedersen 2019). This method entails making the inference that **similarities across individuals and mini-cases** are “empirical fingerprints” of an underlying causal mechanism (Beach and Pedersen 2019, Chapter 9). These hypothesized mechanisms may be the candidate mechanisms identified in the literature by the research team before starting data collection that are supported by the case study evidence, or they may be new mechanisms suggested by the evidence. Identifying new causal mechanisms entails articulating new abstract explanations that are consistent with the observed patterns of evidence. The research team will also use the coded data to identify hypothesized **individual and aggregate-level moderating factors** by identifying differences in the presence or absence of empirical fingerprints of hypothesized mechanisms across mini-cases. When developing these hypothesized mechanisms and moderating factors, each member of the research team working on the case study will once again use informal Bayesian reasoning to articulate natural language probabilities. For mechanisms that were identified before fieldwork, these will be posterior probabilities, updating the prior given the certainty, uniqueness, and validity of the observed evidence ($p(h|e)$). For new mechanisms and moderating factors, these informal probabilities will inform the priors in future rounds of testing the middle-range theory and will be based on the level of confidence in the hypotheses based on the observed evidence.

As in data collection, one potential limitation/risk during this analysis phase is that the dual research/programming role of Mercy Corps will introduce cognitive or social biases that lead research teams to overlook or ignore interpretations and hypotheses that cast Mercy Corps or its programs in an unfavorable light (or which contradict the prior beliefs of key actors within Mercy Corps). As in the data collection phase of the study, this limitation will be addressed by keeping program and policy teams separate from the initial stages of qualitative data coding and identification of hypothesized mechanisms and mediators (and likelihoods about those hypotheses).

Once an initial coding and hypotheses are drafted by the research team members working on each case study, the team will consult Mercy Corps program, policy, and M&E staff in each country, who will articulate their own natural language probabilities for the identified mediators and mechanisms. This approach will allow for incorporation of the tacit technical and political knowledge held by team members, while also ensuring that potential interpretations and hypotheses are not prematurely discounted.

In addition, Mercy Corps program, policy, and M&E team members in each case study context will join the research team in a participatory **stakeholder analysis** to identify deeper political economy factors related to eventual dissemination and use of the middle-range theory and evidence developed in this study (Bryson 2004, Levy and Walton 2013). This analysis will use the coded data and narrative write-ups of the case study and mini cases to update the map of stakeholder interests and power for each case, revising hypothesized positions of actors identified before the analysis, and adding new actors and coalitions surfaced during the data collection and analysis. This stakeholder analysis will be used to inform the final dissemination and evidence use plan for each country, and may also be incorporated into the middle-range theory, if the evidence suggests a hypothesized moderating factor in which the

interests of key stakeholders shape the implementation and effectiveness of governance training interventions.

B.3 Qualitative Analysis Across Cases, Articulation of Working Middle-Range Theory, and Implications for Theory Testing and Evidence Use

To articulate additional hypotheses about context-level moderating factors and scope conditions that are necessary to generate a working middle-range theory, we will compare the working theory developed in each individual case study context to one another. This comparative analysis will also allow us to refine the set of mechanisms and moderators identified in each case and to formulate each hypothesis in abstract terms that can be used in future hypothesis testing (Beach and Pedersen 2019).

As with the within-country process-tracing analysis, we will use informal Bayesian reasoning to develop hypotheses about higher-order moderating factors and scope conditions. To do this, the whole research team will collectively examine the working theories and natural language probabilities from each case study side by side. When hypothesized mechanisms and moderators (and related evidence) are consistent across contexts, the related probability will be adjusted accordingly. When identified mechanisms/moderators differ across context, the team will develop potential explanations for the differing factors, probing the case study analysis for country or program-level factors that may interact with the hypothesized mechanism or may prevent the mechanism from operating in the context (Beach and Pedersen 2019). In addition, the team will examine whether differences in the contents and structure of the working theory across cases are due to mis-specified mechanisms or weaknesses in the qualitative evidence that informed the mechanism. Where necessary to interpret cross-case differences, the team will revisit coded data or seek targeted additional qualitative data sources that can help to refine the definition of concepts or the hypothesized mechanisms, moderators, or scope condition. Each research team member will use this process of discussion and investigation to revise and articulate the natural language probabilities expressing their certainty/uncertainty about each element of the middle-range theory. This process of deliberation and iteration will continue until there is consensus about an aggregate set of natural language probabilities for a unified set of concepts, causal mechanisms, moderating factors, and scope conditions explaining how and where governance trainings shape behaviors and outcomes.

This set of elements will jointly comprise the unified middle-range theory that is the core intellectual output of this study. After coming to a consensus on the working middle-range theory, the research team will hold another joint workshop with key policy, program, and M&E team members from the Jordan and Myanmar offices and Mercy Corps' headquarters offices. As with the in-country workshops, this participatory session will give a broader set of Mercy Corps' stakeholders to assess and input into the working theory and natural language priors, and to add their own contextual, substantive, and political knowledge and interpretations into the middle-range theory (Hoppe 1999). These discussions will also be used to review and finalize the stakeholder analyses and evidence use plans for each country and for global stakeholders.

Once the middle-range theory and stakeholder analyses are finalized, the research team will work with Mercy Corps communications teams to develop synthesized narrative and visual presentations of the middle-range theory. Given that the middle-range theory will have a variety of academic, practitioner, and policy audiences, these synthesized articulations of the theory will vary with respect to level of detail and the type of language used to present the theory's core hypotheses and the evidence from the

two case studies. In each of the final deliverables, we will communicate the natural language probabilities associated with each element of the middle range theory. These probabilities will be used to help identify which hypotheses are the highest priority for future theory-testing as part of future program implementation and evaluation. For these hypotheses, we will also suggest the implications of the middle-range theory for the mix of methods that can be used for future theory testing and the types of contexts and programs that should be the focus of future testing. In addition, the articulation of the natural language probabilities for each element of the theory will be used to identify the mechanisms and moderators that are under-developed and to suggest potential methods and case studies that should be the focus of future theory-building research on governance training interventions.

Annex C: Works Cited

- Acemoglu, Daron, and James Robinson. 2008. "Persistence of Power, Elites, and Institutions." *American Economic Review* 98(1): 267–93.
- Agarwal, Sanjay, and Warren A. Van Wicklin III. 2012. "How, When and Why to Use Demand-Side Governance Approaches in Projects."
- Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. Cambridge, UK: Cambridge University Press.
- Andrews, Matt, Lant Pritchett, and Michael J. V. Woolcock. 2017. *Building State Capability: Evidence, Analysis, Action*. Oxford ; New York, NY: Oxford University Press.
- Baekgaard, Martin et al. 2019. "The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians." *British Journal of Political Science* 49(3): 1117–1140.
- Bain, Katherine, David Booth, and Leni Wild. 2016. "Doing Development Differently at the World Bank: Updating the Plumbing to Fit the Architecture." *London: Overseas Development Institute*.
- Baldwin, Kate, Shylock Muyengwa, and Eric Mvukiyehe. 2017. *Reforming Village-Level Governance via Horizontal Pressure: Evidence from an Experiment in Zimbabwe*. The World Bank.
- Baralt, Melissa. 2012. "Coding Qualitative Data." *Research methods in second language acquisition*: 222.
- Barron, Patrick, Rachael Diprose, and Michael JV Woolcock. 2011. *Contesting Development: Participatory Projects and Local Conflict Dynamics in Indonesia*. Yale University Press.
- Beach, Derek, and Rasmus Brun Pedersen. 2011. "What Is Process-Tracing Actually Tracing? The Three Variants of Process Tracing Methods and Their Uses and Limitations." *The three variants of process tracing methods and their uses and limitations*.
- . 2019. *Process-Tracing Methods: Foundations and Guidelines*. University of Michigan Press.
- Befani, Barbara, and Gavin Stedman-Bryce. 2017. "Process Tracing and Bayesian Updating for Impact Evaluation." *Evaluation* 23(1): 42–60.
- Belotto, Michael J. 2018. "Data Analysis Methods for Qualitative Research: Managing the Challenges of Coding, Interrater Reliability, and Thematic Analysis." *The Qualitative Report* 23(11): 2622–2633.
- Bengtsson, Bo, and Hannu Ruonavaara. 2017. "Comparative Process Tracing: Making Historical Comparison Structured and Focused." *Philosophy of the Social Sciences* 47(1): 44–66.

- Bennett, Andrew, Tasha Fairfield, and Hillel David Soifer. 2019. "Comparative Methods and Process Tracing." *American Political Science Association Organized Section for Qualitative and Multi-Method Research, Qualitative Transparency Deliberations, Working Group Final Reports, Report III*.
- Bicchieri, Cristina. 2016. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. New York, NY: Oxford University Press.
- Blair, Erik. 2015. "A Reflexive Exploration of Two Qualitative Data Coding Techniques." *Journal of Methods and Measurement in the Social Sciences* 6(1): 14–29.
- Blattman, Christopher, Alexandra C. Hartman, and Robert A. Blair. 2014. "How to Promote Order and Property Rights under Weak Rule of Law? An Experiment in Changing Dispute Resolution Behavior through Community Education." *American Political Science Review*: 100–120.
- Boesen, Nils. 2007. "'Governance and Accountability: How Do the Formal and Informal Interplay and Change?'" 2007)'*Informal Institutions, how social norms help or hinder development*'. Paris, OECD Development Centre: p: 83–98.
- Booth, David. 2015. "Thinking and Working Politically." *Development*.
- Booth, David, Daniel Harris, and Leni Wild. 2016. "From Political Economy Analysis to Doing Development Differently: A Learning Experience." *London: ODI*.
- Bryson, John M. 2004. "What to Do When Stakeholders Matter: Stakeholder Identification and Analysis Techniques." *Public management review* 6(1): 21–53.
- Burnard, Philip. 1991. "A Method of Analysing Interview Transcripts in Qualitative Research." *Nurse education today* 11(6): 461–466.
- Cartwright, N. 2020. Using middle-level theory to improve programme and evaluation design. CEDIL Methods Brief. Oxford: CEDIL.
- Cartwright, N., Charlton, Lucy, Juden, Matt, Munslow, Tamlyn & Williams, Richard Beadon. 2020. Making predictions of programme success more reliable. CEDIL Methods Working Paper. Oxford: CEDIL. cedilprogramme.org/wp-content/uploads/2020/10/CEDIL-methods-WP-N.-Cartwright-O
- Casey, Katherine. 2018. "Radical Decentralization: Does Community-Driven Development Work?" *Annual Review of Economics* 10: 139–163.
- CEDIL. 2020. "Guidelines for Producing Research Design PAper."
- Checkel, Jeffrey T. 2006. "Tracing Causal Mechanisms." *International Studies Review* 8(2): 362–370.

- Davey, Calum et al. 2018. "Designing Evaluations to Provide Evidence to Inform Action in New Settings." *CEDIL Inception Paper* (2).
- Diefenbach, Thomas. 2009. "Are Case Studies More than Sophisticated Storytelling?: Methodological Problems of Qualitative Empirical Research Mainly Based on Semi-Structured Interviews." *Quality & Quantity* 43(6): 875.
- Elo, Satu, and Helvi Kyngäs. 2008. "The Qualitative Content Analysis Process." *Journal of Advanced Nursing* 62(1): 107–115.
- Fairfield, Tasha, and Andrew Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." *Political Analysis* 25(3): 363–380.
- . 2019. "A Dialogue with the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." *Perspectives on Politics* 17(1): 154–167.
- Finkel, Steven E., Jeremy Horowitz, and Reynaldo T. Rojo-Mendoza. 2012. "Civic Education and Democratic Backsliding in the Wake of Kenya's Post-2007 Election Violence." *The Journal of politics* 74(1): 52–65.
- Finkel, Steven E., and Amy Erica Smith. 2011. "Civic Education, Political Discussion, and the Social Transmission of Democratic Knowledge and Values in a New Democracy: Kenya 2002." *American Journal of Political Science* 55(2): 417–435.
- Fritz, Verena. 2017. "Doing Development Differently: Understanding the Landscape and Implications of New Approaches to Governance and Public-Sector Reforms." In *Transformation, Politics and Implementation*, JSTOR, 75–98.
- Gentles, Stephen J., Cathy Charles, Jenny Ploeg, and K. Ann McKibbin. 2015. "Sampling in Qualitative Research: Insights from an Overview of the Methods Literature." *The qualitative report* 20(11): 1772–1789.
- George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Illustrated Edition. Cambridge, Mass: The MIT Press.
- Gerring, John. 2004. "What Is a Case Study and What Is It Good For?" *American political science review* 98(2): 341–354.
- . 2007. *Case Study Research: Principles and Practices*. Cambridge University Press.
- Gerring, John, and Lee Cojocar. 2016. "Selecting Cases for Intensive Analysis: A Diversity of Goals and Methods." *Sociological Methods & Research* 45(3): 392–423.
- Gillespie, Andra, and Melissa R. Michelson. 2011. "Participant Observation and the Political Scientist: Possibilities, Priorities, and Practicalities." *PS: Political Science and Politics* 44(2): 261–265.

- Guala, Francesco. 2010. "Extrapolation, Analogy, and Comparative Process Tracing." *Philosophy of Science* 77(5): 1070–1082.
- Gulzar, Saad, and Muhammad Yasir Khan. 2018. *Motivating Political Candidacy and Performance: Experimental Evidence from Pakistan*. unpublished manuscript, Stanford University.
- Gurung, Sanjay, Allison Shean, Carrie O'Neil, and Karen Barnett. 2017. *CATALYSE: Communities Acting Together*. Portland, OR: Mercy Corps. A Governance in Action Guide.
- Guy, David. 2016. "Aid Workers Talk Endlessly About Capacity Building, But What Does It Really Mean?" *The Guardian* 10.
- Hakiman, Kamran, and Ryan Sheely. 2020. *Pathways to Good Governance*. Washington, D.C.: Mercy Corps.
- Hall, Peter A., and Rosemary C. R. Taylor. 1996. "Political Science and the Three New Institutionalisms." *Political Studies* 44(5): 936–57.
- Haruna, Peter F., and Lawrence A. Kannae. 2013. "Connecting Good Governance Principles to the Public Affairs Curriculum: The Case of Ghana Institute of Management and Public Administration." *Journal of Public Affairs Education* 19(3): 493–514.
- Helmke, Gretchen, and Steven Levitsky. 2003. *Informal Institutions and Comparative Politics: A Research Agenda*. Cambridge Univ Press.
<http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=266165> (March 26, 2013).
- Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini. 2019. *How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities*. National Bureau of Economic Research.
- Honig, Dan. 2018. *Navigation by Judgment: Why and When Top Down Management of Foreign Aid Doesn't Work*. Oxford, New York: Oxford University Press.
- Honig, Dan, and Nilima Gulrajani. 2018. "Making Good on Donors' Desire to Do Development Differently." *Third World Quarterly* 39(1): 68–84.
- Hoppe, Robert. 1999. "Policy Analysis, Science and Politics: From 'Speaking Truth to Power' to 'Making Sense Together.'" *Science and public policy* 26(3): 201–10.
- Hsieh, Hsiu-Fang, and Sarah E. Shannon. 2005. "Three Approaches to Qualitative Content Analysis." *Qualitative health research* 15(9): 1277–1288.
- Independent Commission for Aid Impact. 2017. *DFID's Governance Work in Nepal and Uganda, a Performance Review Approach Paper*. <https://icai.independent.gov.uk/wp-content/uploads/ICAI-Governance-Approach-Paper-Final-1.pdf>.

- Ivanoff, Synneve Dahlin, and John Hultberg. 2006. "Understanding the Multiple Realities of Everyday Life: Basic Assumptions in Focus-Group Methodology." *Scandinavian journal of occupational therapy* 13(2): 125–132.
- Jamshed, Shazia. 2014. "Qualitative Research Method-Interviewing and Observation." *Journal of basic and clinical pharmacy* 5(4): 87.
- Joshi, Anuradha, and Mick Moore. 2004. "Institutionalised Co-Production: Unorthodox Public Service Delivery in Challenging Environments." *Journal of Development Studies* 40(4): 31–49.
- Justino, Patricia. 2019. "Governance Interventions in Conflict-Affected Countries." *The Journal of Development Studies* 55(7): 1364–1378.
- Kaidesoja, Tuukka. 2019. "A Dynamic and Multifunctional Account of Middle-Range Theories." *The British journal of sociology* 70(4): 1469–1489.
- Kaidesoja, Tuukka Juhani. 2018. "Building Middle-Range Theories from Case Studies." *Studies in History and Philosophy of Science Part A*.
- Kallio, Hanna, Anna-Maija Pietilä, Martin Johnson, and Mari Kangasniemi. 2016. "Systematic Methodological Review: Developing a Framework for a Qualitative Semi-Structured Interview Guide." *Journal of advanced nursing* 72(12): 2954–2965.
- Kashwan, Prakash, Lauren M. MacLean, and Gustavo A. García-López. 2018. "Rethinking Power and Institutions in the Shadows of Neoliberalism." *World Development*.
- Kelsall, Tim. 2016. "Thinking and Working with Political Settlements." *London: Overseas Development Institute*.
- Khan, Mushtaq. 2010. "Political Settlements and the Governance of Growth-Enhancing Institutions."
- Levy, Brian. 2014. *Working with the Grain: Integrating Governance and Growth in Development Strategies*. Oxford: Oxford University Press.
- Levy, Brian, and Michael Walton. 2013. *Institutions, Incentives and Service Provision: Bringing Politics Back In*. Manchester: Effective States and Inclusive Development Research Center, The University of Manchester. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2386655.
- Liamputtong, Pranee. 2011. *Focus Group Methodology: Principle and Practice*. Sage Publications.
- Longhurst, Robyn. 2003. "Semi-Structured Interviews and Focus Groups." *Key methods in geography* 3(2): 143–156.
- Maher, Carmel, Mark Hadfield, Maggie Hutchings, and Adam de Eyto. 2018. "Ensuring Rigor in Qualitative Data Analysis: A Design Research Approach to Coding Combining NVivo with

Traditional Material Methods.” *International Journal of Qualitative Methods* 17(1): 1609406918786362.

Mahoney, James. 2016. “Mechanisms, Bayesianism, and Process Tracing.” *New Political Economy* 21(5): 493–499.

Mahoney, James, and Kathleen Thelen. 2009. *Explaining Institutional Change: Ambiguity, Agency, and Power*. 1st ed. Cambridge University Press.

Menocal, Alina Rocha, and Pilar Domingo. 2018. “Renewing Representation: The Role of International Assistance in Supporting Democratic Revitalisation.”

Moehler, Devra C. 2010. “Democracy, Governance, and Randomized Development Assistance.” *The Annals of the American Academy of Political and Social Science* 628(1): 30–46.

Mosley, Layna, ed. 2013. *Interview Research in Political Science*. Illustrated Edition. Ithaca: Cornell University Press.

Newcomer, Kathryn E., Harry P. Hatry, and Joseph S. Wholey. 2015. “Conducting Semi-Structured Interviews.” *Handbook of practical program evaluation* 492.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge, UK: Cambridge University Press.

North, Douglass C., John Joseph Wallis, and Barry R. Weingast. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge University Press.

North, Douglass C., and Barry Weingast. 1989. “Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England.” *The journal of economic history* 49(4): 803–832.

Nyumba, Tobias O., Kerrie Wilson, Christina J. Derrick, and Nibedita Mukherjee. 2018. “The Use of Focus Group Discussion Methodology: Insights from Two Decades of Application in Conservation.” *Methods in Ecology and evolution* 9(1): 20–32.

Ostrom, Elinor. 2005. *Understanding Institutional Diversity*. Princeton: Princeton University Press.

Paluck, Elizabeth Levy, and Donald P. Green. 2009. “Deference, Dissent, and Dispute Resolution: An Experimental Intervention Using Mass Media to Change Norms and Behavior in Rwanda.” *American Political Science Review* 103(4): 622–44.

Parker, Andrew, and Jonathan Tritter. 2006. “Focus Group Method and Methodology: Current Practice and Recent Debate.” *International Journal of Research & Method in Education* 29(1): 23–37.

- Phillips, D. et al. 2017. *State-Society Relations in Low-and Middle-Income Countries: An Evidence Gap Map*. London: International Initiative for Impact Evaluation.
- Pierson, Paul. 2000. "The Limits of Design: Explaining Institutional Origins and Change." *Governance* 13(4): 475–499.
- Pritchett, L. 2002. "It Pays to Be Ignorant: A Simple Political Economy of Rigorous Program Evaluation." *Journal of Economic Policy Reform* 5(4): 251–69.
- Pritchett, Lant, Salimah Samji, and Jeffrey S. Hammer. 2013. *It's All about MeE: Using Structured Experiential Learning ('e') to Crawl the Design Space*. Washington, DC: Center for Global Development. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2248785.
- Pritchett, Lant, Kunal Sen, and Eric Werker. 2017. *Deals and Development: The Political Dynamics of Growth Episodes*. Oxford University Press.
- Ravanilla, Nico. 2016. *Attracting Good People into Public Service: Evidence from a Field Experiment in the Philippines*. Working Paper.
- Richards, Lyn, and Janice M. Morse. 2012. *Readme First for a User's Guide to Qualitative Methods*. Sage.
- Risse, Thomas, ed. 2011. *Governance without a State?: Policies and Politics in Areas of Limited Statehood*. New York: Columbia University Press.
- Saldaña, Johnny. 2015. *The Coding Manual for Qualitative Researchers*. Sage.
- Schatz, Edward. 2013. *Political Ethnography: What Immersion Contributes to the Study of Power*. University of Chicago Press.
- Seawright, Jason, and John Gerring. 2008a. "Case Selection Techniques in Case Study Research A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61(2): 294–308.
- . 2008b. "Case Selection Techniques in Case Study Research A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61(2): 294–308.
- Shah, Alpa. 2017. "Ethnography? Participant Observation, a Potentially Revolutionary Praxis." *HAU: Journal of Ethnographic Theory* 7(1): 45–59.
- Stein, Danielle, and Craig Valters. 2012. *Understanding Theory of Change in International Development*. London: London School of Economics. Justice and Security Research Programme Paper. <http://www.lse.ac.uk/internationaldevelopment/research/jsrp/downloads/jsrp1.steinalterspn.pdf>.

- USAID. 2019. *Agency Financial Report Fiscal Year 2019: Promoting a Path to Self-Reliance and Resilience*. Washington, DC: U.S. Agency for International Development.
https://www.usaid.gov/sites/default/files/documents/1868/USAIDFY2019AFR_508R.pdf.
- Van Doeveren, Veerle. 2011. "Rethinking Good Governance: Identifying Common Principles." *Public Integrity* 13(4): 301–318.
- Waddington, Hugh, Juliette Finetti, Marie Gaarder, and Ada Sonnenfeld. 2019. *Does Incorporating Participation and Accountability Improve Development Outcomes? Meta-Analysis and Framework Synthesis*. London: International Initiative for Impact Evaluation.
- Waldner, David. 2012. "Process Tracing and Causal Mechanisms." *The oxford handbook of philosophy of social science*: 65–84.
- Wessells, Michael G. et al. 2012. "The Disconnect between Community-Based Child Protection Mechanisms and the Formal Child Protection System in Rural Sierra Leone: Challenges to Building an Effective National Child Protection System." *Vulnerable Children and Youth Studies* 7(3): 211–227.
- Wild, Leni, David Booth, and Craig Valters. 2017. "Putting Theory into Practice: How DFID Is Doing Development Differently." *London: ODI*.
- White, Howard, Radhika Menon, and Hugh Waddington. 2018. "Community-Driven Development: Does It Build Social Cohesion or Infrastructure? A Mixed-Method Evidence Synthesis."
- Woolcock, Michael. 2013. "Using Case Studies to Explore the External Validity of 'Complex' Development Interventions." *Evaluation* 19(3): 229–248.
- Yanow, Dvora. 2003. "Interpretive Empirical Political Science: What Makes This Not a Subfield of Qualitative Methods." *Qualitative Methods* 1(2): 9–13.
- Yin, Robert K. 2015. *Qualitative Research from Start to Finish*. Guilford publications.
- Zaks, Sherry. 2017. "Relationships Among Rivals (RAR): A Framework for Analyzing Contending Hypotheses in Process Tracing." *Political Analysis* 25(3): 344–362.
- . "Updating Bayesian (s): A Critical Evaluation of Bayesian Process Tracing." *Political Analysis*: 1–17.
- Zlotnick, Jack. 1972. "Bayes' Theorem for Intelligence Analysis." *Studies in Intelligence* 16(2): 43–52.



www.CEDILprogramme.org